What is claimed is:

1. A method of speaker normalization comprising:

a feature parameter extracting step of segmenting an input speech utterance into a constant time length to have frames and extracting acoustic feature parameter of each of the frames;

a frequency converting step of frequency-converting the acoustic feature parameter by using plural frequency conversion coefficients previously defined;

a step of using all the combinations of plural post-conversion feature parameters obtained by the frequency conversion and at least one standard phonemic model, to compute plural similarities or distances between the post-conversion feature parameters of each of the frames and the standard phonemic model;

a step of deciding a frequency converting condition for normalizing the input utterance by using the plural similarities or distances; and

a step of normalizing the input utterance by using the frequency converting condition.

2. A method according to claim 1, wherein the step of deciding a frequency converting condition has a step of mutually comparing between the plural similarities or distances included in an input frame constituted by the frame, a step of selecting for each frame a maximum likelihood, combination of

a phoneme and a frequency conversion coefficient by using a result of comparison, and a step of cumulating the frequency of the frequency conversion coefficient in a maximum likelihood over plural frames and deciding a frequency conversion coefficients in highest frequency as a frequency converting condition.

3. A method according to claim 1, wherein the step of deciding a frequency converting condition has a step of mutually comparing between the plural similarities or distances included in an input frame constituted by the frame, a step of selecting a set of a phoneme of the standard phonemic model and a frequency conversion coefficient that provides a result of maximum likelihood, and a step of deciding the selected frequency conversion coefficient as a frequency converting condition of the frame.

4. A method according to claim 1, wherein the step of computing a similarity or distance further includes a step of computing, for each frame, a ratio in similarity or distance of the phoneme as a weight by using the frame-based acoustic feature parameter of the frame and the standard phonemic model, the step of deciding a frequency converting condition being a step to decide the frequency converting condition by using the weight.

5. A method according to claim 4, wherein the step of computing the similarity or distanc of the phoneme as a weight

includes a step of selecting for each frame a frequency conversion coefficient in a maximum likelihood with respect to all the phonemes of the standard phonemic model, a step of deciding a phoneme based frequency converting condition for all the phonemes, on all the phonemes of the standard phonemic model, from a result of cumulating phoneme by phoneme the frequency converting condition in a maximum likelihood over a plural frames, and a step of using the phoneme-based frequency converting condition and the similarity or distance, to decide for each frame a weight for the phoneme-based frequency converting condition, wherein the step of deciding a frequency converting condition decides a frequency converting condition for the frame by using the weight on the phoneme-based frequency converting condition.

6. A method according to claim 1, wherein, said step of deciding frequency converting condition employs at least vowels in comparing similarities or distances.

7. A method according to claim 1, wherein, said step of deciding frequency converting condition employs only vowels in comparing similarities or distances.

8. An apparatus for speech recognition comprising:

a feature parameter extracting section for segmenting an input speech utterance into a constant time length to have frames and extracting acoustic feature parameter of each frames;

a frequency conversing section for frequency-converting the acoustic feature parameter by using plural frequency conversion coefficients previously defined;

a similarity or distance computing section for using all combinations of plural post-conversion feature parameters obtained by the frequency conversion and at least one standard phonemic model, to compute plural similarities or distances between the post-conversion feature parameter of each frames and the standard phonemic model;

a frequency converting condition deciding section for deciding a frequency converting condition for normalizing the input utterance by using the plural similarities or distances; and

a speech-recognition processing section for recognizing a speech by using the input utterance and a subject-of-recognition acoustic model;

whereby the input utterance is normalized by using the determined frequency converting condition thereby effecting speech recognition.

9. An apparatus according to claim 8, wherein the frequency converting condition deciding section mutually compares between the plural similarities or distances included in an input frame constituted by the frame, selects for each frame a maximum likelihood of combination of a phoneme and a frequency conversion coefficient by using a result of

comparis n, and cumulates th fr quency of th frequency conversion coefficients in the maximum likeliho d over the plural frames and decides a frequency conversion coefficient highest in frequency as a frequency converting condition.

10. An apparatus according to claim 8, wherein the frequency converting condition deciding section mutually compares between the plurality of similarity degrees or distances included in an input frame constituted by the input frame, selects a combination of a phoneme of the standard phonemic model and a frequency conversion coefficient that provides a result of maximum likelihood, and decides the selected frequency conversion coefficient as a frequency converting condition of the frame.

11. An apparatus according to claim 8, wherein the similarity or distance computing section computes, for each frame, a ratio in similarity or distance of the phoneme as a weight by using the frame-based acoustic feature parameter and the standard phonemic model, the frequency converting condition deciding section deciding the frequency converting condition by using the weight.

12. An apparatus according to claim 11, wherein the similarity or distance computing section selects for each frame a frequency conversion coefficient in a maximum likelihood with respect to all the phonemes of the standard phonemic model, decides a ph neme-based frequency conv rting condition for all

the phonemes, on all the phonemes of the standard phonemic model from a result of cumulating phoneme by phoneme the frequency converting condition in a maximum likelihood over plural frames, and uses the phoneme-based frequency converting condition and the similarity degree or distance, to decide a weight for the phoneme-based frequency converting condition for each frame, wherein the frequency converting condition deciding section decides a frequency converting condition for the frame by reflecting the weight on the phoneme-based frequency converting condition.

13. An apparatus according to claim 8, wherein said frequency converting condition deciding section employs at least vowels in comparing similarities or distances.

14. An apparatus according to claim 8, wherein said frequency converting condition deciding section employs only vowels in comparing similarities or distances.

15. An apparatus according to claim 8, comprising a frequency converting condition process display section for displaying, for a user, intermediate data obtained by an internal process of the frequency converting condition deciding section.